

# **Document made available under the Patent Cooperation Treaty (PCT)**

International application number: PCT/US2004/027968

International filing date: 27 August 2004 (27.08.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US

Number: 60/498,393

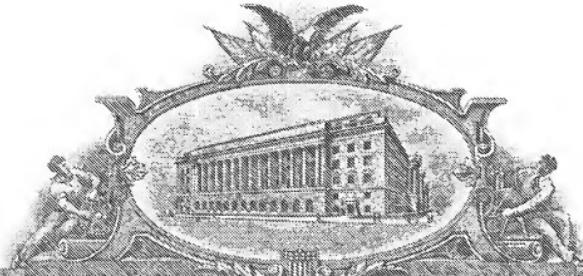
Filing date: 27 August 2003 (27.08.2003)

Date of receipt at the International Bureau: 26 February 2007 (26.02.2007)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland  
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse



# THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

*February 15, 2007*

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: 60/498,393

FILING DATE: *August 27, 2003*

RELATED PCT APPLICATION NUMBER: PCT/US04/27968

THE COUNTRY CODE AND NUMBER OF YOUR PRIORITY APPLICATION, TO BE USED FOR FILING ABROAD UNDER THE PARIS CONVENTION, IS US60/498,393



Certified by

Under Secretary of Commerce  
for Intellectual Property  
and Director of the United States  
Patent and Trademark Office

**Abstract**

This invention provides methods and compositions for treating hyperproliferative conditions such as cancer using reagents relating to the GP115 gene.

## INVENTOR INFORMATION

Inventor One Given Name:: Ronan C  
Family Name:: O'Hagan  
Postal Address Line One:: 15 Garrison Road, Apt. 2  
City:: Bookline  
State or Province:: Massachusetts  
Postal or Zip Code:: 02445  
Citizenship Country:: Canada and Ireland  
Inventor Two Given Name:: Karuppiyah  
Family Name:: Kannan  
Postal Address Line One:: 170 Gore Street, #107  
City:: Cambridge  
State or Province:: Massachusetts  
Postal or Zip Code:: 02141  
Citizenship Country:: India  
Inventor Three Given Name:: David  
Family Name:: Bailey  
Postal Address Line One:: 68 Hobart Street, #2  
City:: Brighton  
State or Province:: Massachusetts  
Postal or Zip Code:: 02135  
Citizenship Country:: British

## CORRESPONDENCE INFORMATION

Correspondence Customer Number:: 1473  
Fax One:: 212.596.9090  
Electronic Mail One:: zli@fishneave.com

## APPLICATION INFORMATION

Title Line One:: GP115: METHODS AND COMPOSITIONS FOR TREA  
Title Line Two:: TING CANCER  
Total Drawing Sheets:: 0  
Formal Drawings?:: No  
Application Type:: Provisional  
Docket Number:: GEN/025PROV  
Secrecy Order in Parent Appl.?:: No

## REPRESENTATIVE INFORMATION

Representative Customer Number:: 1473  
Registration Number One:: 38479  
Registration Number Two:: 42800

Source:: PrintEFS Version 1.0.1

**GP115: METHODS AND COMPOSITIONS FOR TREATING CANCER**

**Field of the Invention**

This invention relates generally to the field of molecular biology. More particularly, this invention relates to genes involved in cancer genesis, maintenance, 5 and progression.

**Background of the Invention**

The impact of cancer on our society cannot be understated. Despite the long history of clinical and research efforts directed towards understanding cancer, surprisingly little is known about the genetic lesions responsible for its genesis, 10 progression, and clinical behavior. For example, in the case of melanoma, although many genes have been implicated in the genesis of this disease, only the INK4a, RAS and BRAF genes have been shown to be true etiologic lesions in a formal genetic sense (Chin et al., Genes Devel. 11:2822-34 (1997); Davies et al., Nature 417:949-54 (2002)). Moreover, advanced malignancy represents the phenotypic endpoint of many 15 successive genetic lesions that affect many oncogene and tumor suppressor gene pathways. Lesions that lead to such a condition may therefore differ from those

required to maintain it. Both types of lesions represent rational therapeutic targets in the treatment of cancer.

#### **Summary of the Invention**

This invention is based on the discovery that a gene designated GP115  
5 is involved in hyperproliferative conditions such as cancer. Up-regulation of GP115 contributes to tumorigenesis (i.e., initiation) and tumor development (including maintenance, progression, and/or metastasis) in a mammal (e.g., a mouse, a nonhuman primate, or a human). GP115 encodes Chromosome 17 open reading frame 26 ("C17orf26"; GenBank Accession No. NM\_139177; UniGene ID Hs.3402), and is  
10 located at human chromosome 17q25.1. An exemplary human GP115 protein has a polypeptide sequence of SEQ ID NO:1. As used herein, "GP115" refers to the gene or a protein product of the gene, depending on the context, of any animal species.

This invention features antibodies specific for the GP115 protein. In some embodiments, the antibodies block the binding of the GP115 protein to its  
15 natural ligand, thereby disrupting the tumorigenic signal transmitted by the ligand. As a result, tumorigenesis and/or tumor development in a patient is delayed, prevented or eliminated.

This invention also features modulators of GP115 activities. These modulators include antagonists of GP115 activities that prevent tumor growth *in vitro*.  
20 In some embodiments, the antagonists can also prevent tumorigenesis, tumor development, and tumor recurrence *in vivo*. The antagonists may inhibit either the biological functions or the production of the GP115 protein in cells. The antagonists include, without limitation, the aforementioned blocking antibodies, small molecular compounds (e.g., organic compounds, peptides, and peptide mimetics), antisense  
25 nucleic acids, ribozymes, and interfering RNAs. The antagonists may reduce GP115 activity levels by at least 50% (e.g., at least 60%, 70%, 80%, or 90%).

Also featured in the invention are therapeutic compositions comprising the above modulators. The compositions may comprise or be used in combination with other cancer therapeutics. Exemplary cancer therapeutics are farnesyl transferase

inhibitors, tamoxifen, herceptin, taxol, ST1571, cisplatin, 5-flurouracil, cytoxin, and irradiation, some of which specifically target members of the *ras* tumorigenic pathway.

Embraced within the invention are also transgenic animals whose genomes comprise a GP115-coding sequence. The GP115-coding sequence may be linked to a constitutive promoter or an inducible promoter to allow overexpression of the GP115 protein. The GP115-coding sequence may contain mutations (e.g., dominant negative mutations and activating mutations), to allow, e.g., extinction of expression or activity of the GP115 protein. Featured in this invention are also animals whose endogenous GP115 gene locus (or loci) comprises (1) a null mutation (i.e., heterozygous or homozygous knockout animals), or (2) a regulatory or coding region mutation engineered through technologies such as knockin modifications on a somatic or germline level.

In one embodiment, the invention features a transgenic animal whose genome comprises: (a) an expression construct comprising a GP115 gene linked operably to an inducible promoter, and (b) a genetic mutation that causes the transgenic mouse to have greater susceptibility to cancer than a mouse not comprising the genetic mutation, where expression of the GP115 gene leads to formation of cancer in the transgenic mouse and the cancer regresses when expression of the GP115 gene is reduced. Mutations that render the animal more susceptible to cancer include disabling mutations in a tumor suppressor gene (e.g., INK4a) and activating mutations in an oncogene (e.g., *myc* and *ras*). In a related embodiment, the induction of the GP115 expression occurs in a tissue-specific manner, i.e., the GP115 transgene can be turned on or off only in a particular tissue of the animal; this embodiment allows one to study the development (including maintenance), regression and recurrence of tumor in a selected tissue or organ of the animal as well as the efficacy and tissue toxicity of candidate drugs that target GP115.

This invention further provides methods for identifying compounds useful in treating hyperproliferative conditions such as cancer. One such method

involves contacting the GP115 gene or protein with a candidate compound, and detecting any reduction in the expression of the gene or activity level of the protein, wherein a reduction indicates that the candidate compound is useful in treating cancer.

In another such method, a biomarker for inhibition of GP115 activity is 5 first identified; and the activity of this biomarker in cells or a mammal contacted by a test compound is then determined, wherein an alteration of the biomarker activity relative to uncontacted cells or an uncontacted mammal indicates that the test compound is a potential anti-cancer drug.

In yet another such method, a first molecular profile (e.g., 10 transcriptional, proteomic or genomic) for GP115 activity is first established by, e.g., identifying a plurality of biomarkers whose patterns of expression or biological function alteration are characteristic of inhibition of GP115 activity in cancer cells; a second molecular profile of biomarkers is established for a candidate compound; and then the two profiles are compared, wherein substantial similarity of the two profiles 15 indicates that the test compound is a potential anti-cancer drug. "Substantial similarity" means that the Pearson correlation coefficient of biomarker expression/activity for the two molecular profiles is statistically significant, with a  $p$  value of less than 0.1 (e.g., less than 0.05, 0.02, or 0.01). The non-overlapping portion between the two profiles may represent nonspecific activity of the candidate 20 compound and allow prediction of the potential toxicity of the compound.

Another such method calls for the use of cells containing a GP115 transgene linked operably to an inducible promoter (e.g., cells derived from the above-described inducible mouse model). In this method, a first molecular profile (e.g., transcriptional, proteomic or genomic) for GP115 activity is first established by, e.g., 25 identifying a plurality of biomarkers whose patterns of expression or biological function alteration are characteristics of switching off the GP115 transgene expression; a second molecular profile is established for a candidate compound; and then the two profiles are compared, wherein substantial similarity of the two profiles indicates that the test compound is a potential anti-cancer drug. Again, the non-

overlapping portion between the two profiles may be indicative of the potential toxicity of the compound.

This invention also provides methods of treating various hyperproliferative conditions such as cancers, psoriasis, arteriosclerosis, arthritis and 5 diabetic retinopathy and other disorders related to uncontrolled angiogenesis and/or vasculogenesis. These methods involve administering the modulators of the invention to a mammal (e.g., a mouse, a rat, a nonhuman primate, or a human). Various stages of cancer are treated by these methods, including neoplasia and malignant tumors. Cancers that can be treated by these methods include, without limitation, cancers that 10 have failed other therapies, cancers at various stages of evolution (including recurring, resistant and minimal residual cancers), cancers whose etiology involves *ras* and/or other members of the *ras* signal transduction pathways, and cancers in which GP115 is overexpressed.

This invention also features methods of diagnosing an abnormal 15 hyperproliferative condition (e.g., cancer) in a subject. These methods involve detecting the expression level of the GP115 gene or the activity level of the GP115 protein, wherein an abnormally high level relative to control (e.g., at least about 50%, 100%, 150%, 200%, 250%, or 300% higher) is indicative of an abnormal hyperproliferative condition.

20 Other features and advantages of the invention will be apparent from the following detailed description.

#### **Detailed Description of the Invention**

This invention is based in part on the discovery that a gene designated GP115 is involved in hyperproliferative conditions such as cancer. GP115 was 25 identified by the Mammalian Second Site Supression ("MaSS") screening system described below and in WO 02/079419, the disclosure of which is incorporated herein in its entirety. Up-regulation of GP115 contributes to tumorigenesis and tumor maintenance in a mammal.

The GP115 protein, i.e., Chromosome 17 open reading frame 26 or C17orf26, has a metal ion transporter activity and more specifically, it is a Zinc transporter protein. An exemplary human GP115 protein contains 335 amino acid residues and has the following polypeptide sequence:

5 MLQGHSSVFQ ALLGTFFTNG MTAAGAALVF VFSSGQRRIL DGSLGFAAGV  
MLAASYWSLL APAVEMATSS GGFGFAFAPPV VAVGFTLGA FVYIADLLMP  
HLGAAEDPQT ALALNFGSTL MKKKSDPEGP ALLFPESELS IRIDKSENGE  
AYQRKKAAT GLPEGPAVPV PSRGNLAQPG GSSWRRRIALL ILAITIHNV  
EGLAVGVGFG AIEKTASATF ESARNLAIGI GIQNFPPEGLA VSLPLRGAGF  
10 STWRWFYWQG LSGMVEPLAG VFGAFAVVLA EPILPYALAF AGAMVYVVM  
DDIIPEAQIS GNGKLASWAS ILGFVVMMMSL DVGLG (SEQ ID NO:1; GenBank  
No. NP\_631916)

A coding sequence for this polypeptide is:

ggagccggcg agagagtggc agcgggggct gatggaagtg cagtgggggc  
15 tggagaggc accctactgt atccagcatg ctccaaggcc acagctctg  
gttccaggcc ttgctggggc ccttcttcac ctggggatg acacgacgt  
ggcagctct cgtgttcgtt ttctcttagtg gacagaggcg gatcttagat  
ggaagtcttg gctttgtgc agggtcatg ttggcagctt cctattggc  
tcttctggcc ccagcagttt agatggccac gtcctctggg ggcttcgggt  
20 ccttgcctt cttccctgtg gctgtggct tcacccttgg agcggctttt  
gtctacttgg ctgacacctt gatgcctcac ttgggtgcag cagaagaccc  
ccagacggcc ctggcactga acttcggctc tacgttgatg aagaagaagt  
ctgatcctga gggcccccg ctgctctcc ctgagagtga actttccatc  
cgatagaca agagtgagaa tggtgaggca tatcagagaa agaaggccgc  
25 agccactggc cttccagagg gtccctgcgt ccctgtgcct tctcgaggga  
atctggcaca gcccggccgc agcagctgg a gaggatcgc actgctcatc  
ttggccatca ctatacacaaa cggtccagag ggtctcgctg ttggatgg  
atttggggct atagaaaaga cggcatctgc taccttgag a gtcgcagg  
atttggccat tggaatcggg atccagaatt tccccgaggg cctggctgtc

agccttcct tgcgaggggc aggcttcctcc acctggagag ctttctggta  
tggcagctg a诶cggcatgg tggagccct ggccggggtc ttgggtcc  
ttggcggtgt gctggcttag cccatccctgc cctacgctct gccttgc  
gcggtgcca tggtctacgt ggtcatggac gacatcatcc ccgaagccca  
5 gatcagtggt aatggaaac tggcatccctg ggcctccatc ctgggatttg  
tagtgtatgtat gtcactggac gttggctgg gctaggctg agacgcttc  
gaccggaga aaggccatac gaagaacag cagtggatgg ctttatggg  
acaacaagct tctttcttca cattaaaact tttttcttc ctctttctt  
catctcatttta tcctgattta ctctgattat aatagaacca tttttacttt  
10 gctttgaggg agattttga ttaatgggg aatttttaagg tgcgtatggaa  
atacagatttccatggg ccactgaatg gactctctct tcagtggtat  
tatcaaggaa cttagatca gggaaatctc cacttcggga ccttctatct  
gcctcccaac tcctcaaggt cacctataga agcgagctac caaaagacgt  
ctcttaagca ttttgggtggc ctgtgactc agggcagagt ggcagcaca  
15 cctctcatcc gcccctccctg ctccatcaact gctgagcctc tccccatcta  
gaatgttggaa actggagcat cataaagata gcaagctacc ttccaaggcc  
gagccagcccc agagaggagc atgtcttctt ttaacccccc ctaaggat  
actacatggg agggggacac agaaaaagg aaggaaatttgc gctgtctgg  
cttttttttttttttttttttttttttttttttttttttttttttttttttttttttt  
20 gaggacaact gtgaactcac agtgagccct gtggaaagaa gagacagaca  
gagtgtgggt ttgttcggag gcctctgtgt tcaatggatt ccaggagcaa  
ggccatttgt cgccgtttcc aaatttctta ggcattttt ttgataagtt  
tatagccatc atgtttctaa gagacttggaa gacaccagca aactgttaga  
actcaaactc ttcaattact caaagaagga gccatttcag ttaactcaag  
25 tgaatgaaag agttttggaa tctgtgtgg gtcctccct gttgaccatt  
tggtaactta taatctgaca aaaactcttg agctgcaaca gccttgc  
gagggtcgat gatggggaaag gaagaaggggg atagggaaaag aagaggttaat  
tttacatttc ccctttaaag taaatttttag ccaactcatc attctgaaat  
gtccctataa agaatgagtc gaactagacc agaagccagc ctactcc  
30 ttacatagct tctccaacag ggtagcaat gacctgtcca cttcaaacac

agataaggcc tgccatcctc attggtaaa ggcacacgtg agactttcag  
tgggctctgc tgagaaggaa ggcagcccaag gagtcaggttgc tgccaggatt  
gcattgtcag tgtctgtct cagagttac acattcaatt gcttccaagg  
gtaatctcc tgctgtga atgtatcag accccaaagg ccaaccttgg  
5 gctgggtcta tgtacgttct tccgaaggcac tgatgatcaa aattgaagac  
acattcagag gtttgcattgg tttagatcaa ctggtgtggt gggttgta  
tgtatgtttt attttatgt ctgtatgt agttctacat aatgcaaaatt  
gtgtttctg atggacaaga cctcataact gtgattaata tcaataaaaa  
ggggatgttg tggaaaaaaa aaaaaaaaaa aaaaaaaaaa (SEQ ID NO:2;  
10 GenBank Accession No. NM\_139177)

The open reading frame of the above sequence is nucleotides 78-1085. Other human GP115 coding sequences include Genbank accession numbers AF331643.1, NM\_139177.1, and BC035631.1. GP115 orthologs in other animal species have also been identified. They include GenBank accession numbers  
15 NM\_027216 (*M. musculus*), XM\_213529 (*R. norvegicus*), and NM\_059213 (*C. elegans*).

The GP115 gene is expressed in a variety of tissues, including colon, kidney, lung, heart, myeloid cells, stomach, cervix, uterus, endometrium, B-cells, liver, spleen, melanocyte, marrow, placenta, blood, brain, testis, leukocyte, islets of Langerhans, ascites, osteoarthritic cartilage and breast. The expression of this gene is increased in a broad spectrum of cancer tissue types (e.g., retinoblastoma, hypernephroma, ovarian tumor, squamous cell carcinomas, melanotic melanoma, chronic lymphocytic leukemia, head and neck cancer, leiomyosarcoma, anaplastic oligodendrogloma and metastatic chondrosarcoma) and in cancer cell lines (e.g.,  
20 adenocarcinoma cell line, high MDR cell line, epidermoid carcinoma cell line, cervical carcinoma cell line, lymphoma cell line, hepatocellular carcinoma cell line, ductal carcinoma cell line, human chondrosarcoma cell line, HCT-116, DLD-1, SK-MEL28, A549, MiaPaCa2, LNCaP, FGC, 22Rv1, DU145, MCF-7, PC-3, SW620, HCT-15, and KATO III).

The GP115 gene was found to be involved in development (including maintenance, progression, angiogenesis, and/or metastasis) of cancers, e.g., cancers found in skin (e.g., melanoma), lung, prostate, breast, colorectal, liver, pancreatic, brain, testicular, ovarian, uterine, cervical, kidney, thyroid, bladder, esophageal, and 5 hematological tissues. The cancer-related functions of this gene can be confirmed by, e.g., (1) its overexpression in more than about 10% of human cancers; (2) the inhibition of human cancer cell proliferation and growth in soft agar by RNA interference ("RNAi") of this gene; (3) the ability of this gene to enhance transformation of MEFs by *myc* and *ras*-V12; (4) the ability of this gene to enhance 10 the growth of mouse tumor cell lines in soft agar; and/or (5) the prevention by inhibiting this gene of maintenance or formation of tumors arising *de novo* in a mouse, or tumors derived from human cancer cell lines.

#### I. GP115-RELATED NUCLEIC ACIDS

All the nucleic acid sequences specifically given herein are set forth as 15 sequences of deoxyribonucleotides. It is intended, however, that the given sequences be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine. Further, differences from naturally occurring nucleic acids — e.g., non-native bases, altered internucleoside 20 linkages, post-synthesis modification — can be present throughout the length of the GP115 nucleic acid or can usefully be localized to discrete portions thereof. For example, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and demonstrated utility for targeted gene repair. See, e.g., U.S. Pat. Nos. 5,760,012 and 5,731,181.

25 Polymorphisms such as single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. Additionally, small deletions and insertions, rather than SNPs, are not uncommon in the general population, and often do not alter the function of the protein. Accordingly, this invention provides not only isolated nucleic acids identical in sequence to those described with particularity herein, but also

isolated nucleic acids that are allelic variants of those particularly described nucleic acid sequences. Further, sequence variations can be nonnaturally occurring, i.e., can result from human intervention, as by random or directed mutagenesis.

**A. Nucleic Acids Encoding GP115 Protein or Portions Thereof**

The invention provides isolated nucleic acid molecules that encode the entirety or part (e.g., at least five, seven, or nine contiguous amino acid residues) of the GPC protein, including allelic variants of this protein. As is well known, the genetic code is degenerate and codon choice for optimal expression varies from species to species. Thus, the coding sequences of this invention include degenerate variants of the sequences described herein with particularity. Thus, the isolated polynucleotide comprises a nucleotide sequence encoding SEQ ID NO:1.

These nucleic acids can be used, for example, to express the GP115 protein or specific portions of the protein, either alone or as elements of a fusion protein (e.g., to express epitopic or immunogenic fragments of the GP115 protein).  
These nucleic acids can be used also as probes to hybridize to GP115 nucleic acids and related nucleic acid sequences.

This invention also includes nucleic acids comprising sequences coding for polypeptides containing conservative amino acid substitutions or moderately conservative amino acid substitutions from those polypeptides described with particularity herein. These amino acid substitutions can be due to, e.g., allelic variations, naturally occurring mutations, or manmade mutations.

**B. Cross-Hybridizing Nucleic Acids**

This invention also provides isolated polynucleotides that hybridize to various of the GP115 nucleic acids of this invention. These cross-hybridizing nucleic acids can be used, e.g., as hybridization probes, primers, and/or to drive expression of proteins that are related to GP115 as isoforms and homologs (e.g., paralogs, and orthologs). In some such embodiments, the invention provides an isolated nucleic acid comprising a sequence that hybridizes under high stringency conditions to a

probe comprising a fragment of SEQ ID NO:2 having at least 15, 16, 18, 20, 24, or 25 nucleotides. As used herein, "high stringency conditions" are defined for solution phase hybridization as aqueous hybridization (i.e., free of formamide) in 6X SSC (where 20X SSC contains 3.0 M NaCl and 0.3 M sodium citrate), 1% SDS at 65°C for 5 8-12 hours, followed by two washes in 0.2X SSC, 0.1% SDS at 65°C for 20 minutes. It will be appreciated by the skilled worker that hybridization at 65°C will occur at different rates depending on a number of factors including the length and percent identity of the sequences which are hybridizing.

10 The hybridizing portion of a reference nucleic acid (i.e., target nucleic acid) is typically at least 15 nucleotides in length, and often at least 17 , 20, 25, 30, 35, 40 or 50 nucleotides in length. Cross-hybridizing nucleic acids that hybridize to a larger portion of the reference nucleic acid – for example, to a portion of at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 or more nucleotides, up to and including the entire length of the reference nucleic acid, are also useful.

15 **C. Nucleic Acid Fragments**

Fragments of the above-described nucleic acids are also included in this invention. They can be used as region-specific probes, as amplification primers, regulatory sequences to direct expression of a gene, and/or to direct expression of a GP115 polypeptide fragment (e.g., immunogenic fragment).

20 The nucleic acid probes may comprise a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic kit for identifying cells or tissues (i) that mis-express a GP115 protein (e.g., aberrant splicing, or abnormal mRNA levels), or (ii) that harbor a mutation in the GP115 gene, such as a deletion, an insertion, or a 25 point mutation. Such diagnostic kits preferably include labeled reagents and instructional inserts for their use.

The nucleic acid primers can be used in PCR, primer extension and the like. They can be, e.g., at least 6 nucleotides (e.g., at least 7, 8, 9, or 10) in length.

The primers can hybridize to an exonic sequence of a GP115 gene, for, e.g., amplification of a GP115 mRNA or cDNA. Alternatively, the primers can hybridize to an intronic sequence or an upstream or downstream regulatory sequence of a GP115 gene, to utilize non-transcribed, e.g., regulatory portions of the genomic structure of a GP115 gene. The nucleic acid primers can also be used, for example, to prime single base extension (SBE) for SNP detection (see, e.g., U.S. Pat. No. 6,004,744). Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. See, e.g., Schweitzer et al., *Curr. Opin. Biotechnol.* 12(1):21-7 (2001); U.S. Patent Nos. 5,854,033 and 5,714,320 and international patent publications WO 97/19193 and WO 00/15779. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. See, e.g., Lizardi et al., *Nature Genet.* 19(3):225-32 (1998).

Nucleic acid fragments that encode 5 or more contiguous amino acids (i.e., fragments of 15, 18, 21, 24, or 27 nucleotides or more) are useful in directing the synthesis of peptides that have utility in mapping the epitopes of the GP115 protein. See, e.g., Geysen et al., *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and U.S. Pat. Nos. 4,708,871 and 5,595,915. Such nucleic acid fragments are also useful in directing the synthesis of peptides that have utility as immunogens. See, e.g., Lerner, "Tapping the immunological repertoire to produce antibodies of predetermined specificity," *Nature* 299:592-596 (1982); Shinnick et al., *Annu. Rev. Microbiol.* 37:425-46 (1983); Sutcliffe et al., *Science* 219:660-6 (1983).

Of course, larger fragments having at least 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 or more nucleotides are also useful, and at times preferred, as will be appreciated by the skilled worker.

25           **D. Single Exon Probes**

The invention further provides single exon probes having portions of no more than one exon of the GP115 gene. Such single exon probes have particular utility in identifying and characterizing splice variants. In particular, these probes are useful for identifying and discriminating the expression of distinct isoforms of GP115.

## E. Antisense Reagents

### 1. Antisense Nucleic Acids

Some of the new isolated nucleic acids are antisense polynucleotides that specifically hybridize to GP115 sense polynucleotides. The antisense nucleic acid molecule can be complementary to the entire coding or non-coding region of GP115, but more often is an oligonucleotide that is antisense to only a portion of the coding or non-coding region of GP115 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of GP115 mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 10 35, 40, 45 or 50 nucleotides in length.

The antisense nucleic acids of this invention may, for example, form a stable duplex with its target sequence, or, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. In yet other embodiments, the antisense nucleic acid 15 molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier et al. Nucl. Acids Res 15: 6625-6641 (1987)).

An antisense target sequence is a nucleotide sequence unique to GP115, 20 and can be identified through use of any publicly available sequence database, and/or through use of commercially available sequence comparison programs. Antisense nucleic acids of the invention can then be constructed using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be produced biologically using an expression vector into which a nucleic acid has been 25 inserted in an antisense orientation, i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest.

Alternatively, the antisense nucleic acid can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to

increase the biological stability of the molecule or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids. For example, phosphorothioate derivatives and acridine substituted nucleotides can be used.

5 Phosphorothioate and methylphosphonate antisense oligonucleotides are contemplated for therapeutic use by the invention. The antisense nucleic acid molecule can also comprise a 2'-O-methylribonucleotide (Inoue et al., Nucl. Acids Res 15: 6131-6148 (1987)) or a chimeric RNA -DNA analogue (Inoue et al. FEBS Lett 215: 327-330 (1987)). The antisense oligonucleotides may be further modified by adding poly-L-lysine, transferrin, polylysine, or cholesterol moieties at their 5' end.

10 The antisense nucleic acid molecules of the invention (e.g., oligonucleotides of 10 to 20 nucleotides in length) can be administered to a subject (e.g., a human) or generated *in situ* via an expression vector, such that they bind to cellular RNA and/or genomic DNA encoding a GP115 protein to inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Suppression of GP115 expression at either the transcriptional or translational level is useful to treat certain cancer conditions in patients or to generate cellular or animal models for cancer characterized by aberrant GP115 expression.

20 By way of example, an antisense molecule can be administered by direct injection at a tissue site of a subject. Alternatively, an antisense molecule can be designed to target selected cells (e.g., cancer cells overexpressing GP115) and then administered systemically; for example, the antisense molecule contains a peptide or antibody that specifically binds to a cell surface receptor or antigen expressed on the surface of the selected cell surface. See also discussions below on PNAs. The antisense nucleic acids can also be delivered to target cells using expression vectors 25 encoding thereof. The expression vectors may contain a strong pol II or pol III promoter to ensure that the antisense nucleic acids are expressed at sufficient intracellular concentrations.

## 2. Ribozymes and Catalytic Nucleic Acids

In still another series of embodiments, an antisense nucleic acid of the invention is part of a GP115-specific ribozyme (or, as modified, a "nucleozyme").  
5 Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes such as hammerhead, hairpin, and Group I intron ribozymes can cleave GP115 mRNA transcripts catalytically, thereby inhibiting translation of GP115 mRNA. A ribozyme having specificity for a GP115-encoding nucleic acid can be designed based upon the nucleotide sequence of a GP115  
10 polynucleotide disclosed herein (SEQ ID NO:2). See, e.g., U.S. Patent Nos. 5,116,742; 5,334,711; 5,652,094; and 6,204,027. For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a GP115-encoding mRNA. See, e.g., Cech et al. U.S. Pat. No. 4,987,071; and Cech et al. U.S.  
15 Pat. No. 5,116,742. Alternatively, GP115 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel et al., Science 261:1411-1418 (1993).

In some embodiments, the ribozymes and other antisense reagents of this invention may include appended groups such as peptides (e.g., for targeting host  
20 cell receptors, e.g., the GP115 protein, *in vivo*); agents facilitating transport across the cell membrane (e.g., Letsinger et al., Proc. Natl. Acad. Sci. USA 86:6553-6556 (1989); Lemaitre et al., Proc. Natl. Acad. Sci. USA 84:648-652 (1987); PCT Publication No. W088/09810) or the blood-brain barrier (e.g., PCT Publication No. W089/10134); hybridization-triggered cross-linking agents; and hybridization-  
25 triggered cleavage agents.

Expression of the GP115 gene may be inhibited by targeting nucleotide sequences complementary to the regulatory region of the GP115 (e.g., the GP115 promoter and/or enhancers) to form triple helical structures that prevent transcription of the GP115 gene in target cells. See generally, Helene, Anticancer Drug Des. 6:

569-84 (1991); Helene et al. Ann. N.Y. Acad. Sci. 660:27-36 (1992); and Maher Bioassays 14: 807-15 (1992).

### 3. Peptide Nucleic Acids (PNA)

Some preferred oligonucleotide mimetics, especially those useful for *in vivo* administration, are peptide nucleic acids (PNA). See, e.g., Hyrup et al. Bioorg. Med. Chem. Lett. 4:5-23 (1996). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza-nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al., *supra*; and Perry-O'Keefe et al., Proc. Natl. Acad. Sci. USA 93:14670-675 (1996).

PNAs of GP115 can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigenic agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs of GP115 can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases; or as probes or primers for DNA sequence and hybridization (Hyrup et al., *supra*; and Perry-O'Keefe, *supra*).

In other embodiments, PNAs of GP115 can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of GP115 can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers

of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup et al., *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *supra* and Finn et al., *Nucl. Acids Res.* 24:3357-63 (1996).

5           **4. RNA Interference**

RNA interference (RNAi) is another way to silence the expression of the GP115 gene. RNAi is a sequence-specific posttranscriptional gene silencing mechanism triggered by double-stranded RNA (dsRNA) and causes degradation of mRNAs homologous in sequence to the dsRNA. The mediators of the degradation are 10 21- to 23-nucleotide small interfering RNAs (siRNAs) generated by ribonuclease III cleavage from the longer dsRNAs. Molecules of siRNA typically have 2- to 3-nucleotide 3' overhanging ends resembling the RNase III processing products of long dsRNAs that normally initiate RNAi. When introduced into a cell, they assemble with an endonuclease complex (RNA-induced silencing complex), which then guides target 15 mRNA cleavage. As a consequence of degradation of the targeted mRNA, cells with a specific phenotype of the suppression of the corresponding protein product are obtained (e.g., reduction of tumor size, metastasis, angiogenesis, and growth rates). The small size of siRNAs, compared with traditional antisense molecules, prevents activation of the dsRNA-inducible interferon system present in mammalian cells. This 20 helps avoid the nonspecific phenotypes normally produced by dsRNA larger than 30 base pairs in somatic cells. For review, see, e.g., Elbashir et al., *Methods* 26:199-213 (2002); McManus and Sharp, *Nature Reviews* 3:737-747 (2002); Hannon, *Nature* 418:244-251 (2002); Tuschl, *Nature Biotechnology* 20:446-448 (2002); and Tuschl U.S. Application US2002/0086356 A1.

25           SiRNA oligonucleotides can be designed with a number of software programs, e.g., the OligoEngine siRNA design tool available at <http://www.oligoengine.com>. Preferred siRNAs of this invention range about 19-29 basepairs in length for the double-stranded portion. In some embodiments, the siRNAs are hairpin RNAs having an about 19-29 bp stem and an about 4-34

nucleotide loop. In some embodiments, siRNAs are highly specific for a GP115 target region and may comprise any 19-29 bp fragment of a GP115 mRNA that has at least 1 (e.g., at least 2 or 3) bp mismatch with a nonGP115-related sequence. In some embodiments, the preferred siRNAs do not bind to RNAs having more than 3  
5 mismatches with the target region. The target sequences of exemplary siRNAs for GP115 are:

5'- ctccggctcacgttgatg -3' (SEQ ID NO:3);  
5'- tggtagggcatatcagaga -3' (SEQ ID NO:4); and  
5'- gcctactccttacata-3' (SEQ ID NO:5).

10 These three sequences correspond to nucleotides 422-440, 521-539 and 2239-2257 of SEQ ID NO:2, respectively. SiRNAs that target an RNA region having 10 or more nucleotide overlaps with an aforementioned exemplary target region are also useful.

15 Intracellular transcription of siRNAs can be achieved by cloning the siRNA templates into RNA polymerase III (Pol III) transcription units, which normally encode the small nuclear RNA U6 or the human RNase P RNA H1. Two approaches can be used for expressing siRNA: (1) sense and antisense strands constituting the siRNA duplex are transcribed by individual promoters; or (2) siRNAs are expressed as fold-back stem-loop structures that give rise to siRNAs after  
20 intracellular processing. Inducible promoters can also be used to drive the expression of the siRNA.

By way of example, to target specific regions of a GP115 mRNA, the following oligonucleotides can be used with a primer specific to the U6 small RNA promoter to form double-stranded DNA in a polymerase chain reaction, using a vector  
25 containing this U6 promoter as a template. The PCR product can then be ligated into a vector. Expression of the insert will lead to expression of a short hairpin RNA. The hairpin structure will have inhibitory effects on GP115 expression.

1. GP115 target nucleotides 1-29 (in boldface)

ggaattcaaaaagccccactaccactctccgcgactccccaa**ggagccggcgag**  
**agagtggcagggggc**tatgtatgtgctgccaagc (SEQ ID NO:6)

5

2. GP115 target nucleotides 2664-92 (in boldface)

ggaattcaaaaagatattaatcacagtcatgagatc**cgtccccaaga**cacaagac**ctca**  
**taactgtgattaata**tctatgtatgtgctgccaagc (SEQ ID NO:7)

10 3. GP115 target nucleotides 862-90 (in boldface)

ggaattcaaaaagccccagccagaggctccaccataccacccaa**ggggcatggtg**  
**agccccctggccgggtc**tatgtatgtgctgccaagc (SEQ ID NO:8)

4. GP115 target nucleotides 383-411 (in boldface)

15 ggaattcaaaaaggccgcctaggagtcttctgctgcaccccaaa**gggtgcagcaga**  
**agaccccccagacggccct**atgtatgtgctgccaagc (SEQ ID NO:9)

An siRNA oligonucleotide or its coding sequence can be delivered into a target cell via a variety of methods, including but not limited to, liposome fusion (transposomes), routine nucleic acid transfection methods such as electroporation, 20 calcium phosphate precipitation, and microinjection, and infection by viral vectors.

## F. Exemplary Uses of GP115 Nucleic Acids

### 1. Characterization of Genetic Mutations

The isolated nucleic acids of this invention can be used as hybridization probes to characterize GP115 nucleic acids in both genomic and transcript-derived 25 nucleic acid samples. For example, the probes can be used to detect gross alterations in the GP115 genomic locus, such as deletions, insertions, translocations, and

duplications of the GP115 genomic locus. Methods of detection include fluorescence *in situ* hybridization (FISH) to chromosome spreads, comparative genomic hybridization (CGH), array CGH (e.g., on microarrays containing GP115-coding sequences or BAC comprising GP115-coding sequences), and spectral karyotyping

- 5 (SKY). See, e.g., Andreeff et al. (eds.), *Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications*, John Wiley & Sons (1999) (ISBN: 0471013455). The probes can also be used to assess smaller genomic alterations using, e.g., Southern blot detection of restriction fragment length polymorphisms. The nucleic acid probes can be also used to isolate genomic clones
- 10 that include the nucleic acids of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, amplifications, translocations, and substitutions (e.g., SNPs) at the sequence level. The nucleic acid probes can also be used to isolate GP115 nucleic acids from cDNA libraries, permitting sequence level characterization of GP115 RNA messages, including
- 15 identification of deletions, insertions, truncations — including deletions, insertions, and truncations of exons in alternatively spliced forms — and single nucleotide polymorphisms. Some nucleic acids of this invention can also be used as amplification primers for PCR (e.g., real time PCR) to detect the above-described genomic alterations. Such genomic alterations of the GP115 gene often play a role in
- 20 tumor genesis, maintenance and development, and thus can be used as markers for diagnosis and prognosis of GP115-mediated cancers.

## 2. Quantification of Expression Levels

The nucleic acid probes can be used to measure the representation of GP115 clones in a cDNA library, used as primers for quantitative real time PCR, or 25 otherwise used to measure expression level of the GP115 gene. Measurement of GP115 expression has particular utility in diagnostic assays for cancer-related conditions associated with abnormal GP115 expression. Moreover, differences in the expression levels of the gene before and after a cancer event (e.g., cancer genesis, maintenance, regression, and metastasis) are useful in determining the effect of a

candidate cancer drug, identifying cancer types, designing diagnostics and prognostics, and predicting likely outcome of a cancer therapy.

### 3. Microarrays

- The nucleic acid probes can be bound to a substrate. The substrate can
- 5 be porous (e.g., a membrane comprising nitrocellulose, nylon, or positively-charged derivatized nylon) or solid (e.g., glass, silicon, or plastics). It can be planar or non-planar, unitary or distributed, and the bond can be covalent or noncovalent. In some embodiments, the nucleic acids are bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound
- 10 nucleic acids being separately detectable. At low density, e.g., on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities.
- 15 The microarrays are useful for measuring gene expression in, for example, drug discovery and target validation programs. For instance, where gene expression analysis is used to assess the effectiveness of a candidate cancer drug, a microarray containing a group of cancer-related nucleic acid probes, including GP115 probes, can be used. Decreases in the expression levels of genes that are up-regulated
- 20 in cancer genesis and development (e.g., EGFR, BDNF, HER2/Neu, erb-B2, TGF $\beta$ , RhoC, VEGF-C, KRAS, HRAS and AKT) and/or increases in the expression levels of genes that are typically down-regulated in those events are indicative of the effectiveness of the drug. The microarray may further contain probes for genes that are indicative of toxicity of a drug in humans so that the toxicity of the drug can be
- 25 studied concurrently.

The single exon probes and single exon probe microarrays of the invention have the additional utility of permitting high-throughput detection of splice variants of the nucleic acids of the present invention.

#### 4. Gene Therapy

Since up-regulation of GP115 expression or activity is conducive to tumor formation and/or maintenance, it is desirable to reduce the activity level of GP115 in a cancer patient. One way to achieve this is by delivering into the tumor site 5 an RNA reagent of this invention, including but not limited to, antisense agents and interfering RNAs. But in diseases in which GP115 is down-regulated, it may be desired to insert a functional GP115 gene into the diseased subject by gene therapy.

10 *In vivo* expression of the therapeutic nucleic acids can be driven from a vector — typically a viral vector, often a vector based upon a replication incompetent lentivirus, retrovirus, adenovirus, or adeno-associated virus (AAV). These vector systems can be used either positively or negatively. That is, they can deliver either the GP115 gene in cases where up-regulation of GP115 is desired or inhibitors of GP115 expression such as RNAi in cases where inhibition of GP115 function is desired. *In vivo* expression can be driven from expression control signals endogenous or 15 exogenous (e.g., from a vector) to the nucleic acid.

#### 5. Genetic Alterations

The nucleic acids of this invention can also be used to introduce mutations (e.g., null mutations, dominant negative mutations, dominant acting mutations) into a GP115 locus of an animal via homologous recombination. Such 20 animals (e.g., knock out mice) are useful in delineating the role of GP115 in tumor genesis and development and in facilitating cancer drug development.

Where the genomic region includes transcription regulatory elements, homologous recombination can be used to replace the endogenous regulatory elements with heterologous regulatory elements, i.e., elements not natively associated with the 25 gene in the same manner. This can alter the expression of GP115, both for production of GP115 protein, and for gene therapy. See, e.g., U.S. Pat. Nos. 5,981,214, 6,048,524; and 5,272,071.

Fragments of the polynucleotides of the present invention smaller than those typically used for homologous recombination can also be used for targeted gene correction or alteration, possibly by cellular mechanisms different from those engaged during homologous recombination. See, e.g., U.S. Pat. Nos. 5,945,339, 5,888,983,

- 5 5,871,984, 5,795,972, 5,780,296, 5,760,012, 5,756,325, 5,731,181; and Culver et al., "Correction of chromosomal point mutations in human cells with bifunctional oligonucleotides," *Nature Biotechnol.* 17(10):989-93 (1999); Gamper et al., *Nucl. Acids Res.* 28(21):4332-9 (2000).

## **II. VECTORS AND HOST CELLS**

### **10 A. General Consideration**

This invention provides nucleic acid constructs containing one or more of the isolated nucleic acid molecules of the invention. The vectors can be used to propagate the new nucleic acid molecules in host cells, to shuttle the molecules between host cells derived from disparate organisms, to insert the molecules into host 15 genomes, to express sense or antisense RNA transcripts or interfering RNAs, and/or to express GP115 polypeptides. Typically, the vectors are derived from virus, plasmid, prokaryotic or eukaryotic chromosomal elements, or some combination thereof, and may optionally include at least one origin of replication, at least one site for insertion of heterologous nucleic acid, and at least one selectable marker.

20 This invention further includes host cells — either prokaryotic (bacteria) or eukaryotic (e.g., yeast, insect, plant and animal cells) — containing the new nucleic acid constructs that exist within the cells either episomally or integrated, in whole or in part, into the host cell chromosome. A host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-25 translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, hydroxylation, sulfation, lipidation, and acylation, and it is an aspect of the present invention to provide GP115 proteins with such post-translational modifications.

Exemplary prokaryotic host cells are *E. coli*, *Caulobacter crescentus*, *Streptomyces species*, and *Salmonella typhimurium* cells. Vectors useable in these cells include, without limitation, those related to pBR322 and the pUC plasmids.

- Exemplary yeast host cells are *Saccharomyces cerevisiae*,
- 5      *Schizosaccharomyces pombe*, *Pichia pastoris*, and *Pichia methanolica*. Vectors useable in these host cells are integrative YIp vectors, replicating episomal YEp vectors containing centromere sequences CEN and autonomously replicating sequences ARS.

- Insect cells are often chosen for high efficiency protein expression.
- 10     Exemplary insect host cells are those from *Spodoptera frugiperda* (e.g., Sf9 and Sf21 cell lines, and EXPRESSFTM cells (Protein Sciences Corp., Meriden, CT, USA)), *Drosophila* S2 cells, and *Trichoplusia ni* HIGH FIVE® Cells (Invitrogen, Carlsbad, CA, USA). Where the host cells are *Spodoptera frugiperda* cells, the vector replicative strategy is typically based upon the baculovirus life cycle.
- 15     Exemplary mammalian host cells are COS1 and COS7 cells, NSO cells, Chinese hamster ovary (CHO) cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK, HEK293, WI38, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562, Jurkat cells, BW5147 and any other commercially available human cancer cell lines. Cells with K-ras<sup>G13D</sup>, such as human
- 20     colon cancer cell lines DLD-1 and HCT-116, and revertants thereof having a null mutation in the activated K-ras gene, can also be used. Other useful mammalian cell lines are well known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). As used herein, mammalian host cells also include those in the body of a subject (e.g., a human patient or an animal).

Vectors intended for autonomous extrachromosomal replication in mammalian cells will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells),

the papillomavirus origin, or the EBV origin for long term episomal replication (for use in, e.g., 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of 5 replication functional in mammalian cells, such as the SV40 origin. Useful vectors also include vectors based on viruses such as lentiviruses, adenovirus, adeno-associated virus, vaccinia virus, parvoviruses, herpesviruses, poxviruses, Semliki Forest viruses, and retroviruses.

Plant cells can also be used for expression, with the vector replicon 10 typically derived from a plant virus (e.g., cauliflower mosaic virus (CaMV); tobacco mosaic virus (TMV)) and selectable markers chosen for suitability in plants.

To propagate nucleic acids larger than can readily be accommodated in viral or plasmid vectors, the invention further provides artificial chromosomes — bacterial artificial chromosomes (BACs), yeast artificial chromosomes (YACs), 15 mammalian artificial chromosomes (MACs), and human artificial chromosomes (HACs) — that contain the GP115 nucleic acid of interest.

Vectors of the invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the 20 nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

#### B. Transcription Regulators for Expression Vectors

Expression vectors of the invention often include a variety of other genetic elements operatively linked to the protein-encoding heterologous nucleic acid 25 insert, typically genetic elements that drive and regulate transcription, such as promoters and enhancer elements, those that facilitate RNA processing, such as transcription termination, splicing signals and/or polyadenylation signals, and those that facilitate translation, such as ribosomal consensus sequences. Other transcription control sequences include, e.g., operators, silencers, and the like. Use of such

expression control elements, including those that confer constitutive or inducible expression, and developmental or tissue-regulated expression are well-known in the art.

Constitutively active promoters include, without limitation, a CMV 5 promoter, EF1 $\alpha$ , retroviral LTRs, and SV40 early region.

Inducible promoters useful in this invention include, without limitation, a tetracycline-inducible promoter, a metallothionein promoter, the IPTG/lacI promoter system, the ecdysone promoter system, and the "lox stop lox" system for irreversibly deleting inhibitory sequences for translation or transcription. In one embodiment, a 10 GP115 gene is placed between lox sites, and upon expression of the cre enzyme, the GP115 gene is deleted from the genome so that the GP115 activity is permanently eliminated.

Instead of inducible promoters, the activity of a GP115 protein can also be inducibly switched on or off by fusing the GP115 protein to, e.g., an estrogen 15 receptor polypeptide sequence, where administration of estrogen or an estrogen analog (e.g., hydroxytamoxifen) will allow the correct folding of the GP115 polypeptide into a functional protein.

Tissue-specific promoters that can be used in driving expression of GP115 in animal models include, without limitation: a tyrosinase promoter or a TRP2 20 promoter in the case of melanoma cells and melanocytes; an MMTV or WAP promoter in the case of breast cells and/or cancers; a Villin or FABP promoter in the case of intestinal cells and/or cancers; a RIP promoter in the case of pancreatic beta cells; a Keratin promoter in the case of keratinocytes; a Probasin promoter in the case of prostatic epithelium; a Nestin or GFAP promoter in the case of CNS cells and/or 25 cancers; a Tyrosine Hydroxylase, S100 promoter or neurofilament promoter in the case of neurons; the pancreas-specific promoter described in Edlund et al. Science 230:912-916 (1985); a Clara cell secretory protein promoter in the case of lung cancer; and an Alpha myosin promoter in the case of cardiac cells.

Developmentally regulated promoters may also be selected. They include, without limitation, the murine hox promoters (Kessel and Gruss, *Science* 249:374-379 (1990)) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman, *Genes Dev.* 3:537-546 (1989)).

5       **C. Expression Vectors Encoding Peptide Tags**

Expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Many such tags are known and available. Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides larger than 10 purification and/or identification tags. Useful protein fusions include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as luciferase or those that have a green fluorescent protein (GFP)-like chromophore, and fusions for use in two hybrid selection systems.

For secretion of expressed proteins, a wide variety of vectors are 15 available which include appropriate sequences that encode secretion signals, such as leader peptides. Vectors designed for phage display, yeast display, and mammalian display, for example, target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain.

A wide variety of vectors now exist that fuse proteins encoded by 20 heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its many color-shifted and/or stabilized variants.

For long-term, high-yield recombinant production of the proteins, 25 protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors (preferably having selectable markers), followed by selection for integrants.

### **III. GP115 PROTEINS, POLYPEPTIDES AND FRAGMENTS**

The present invention provides GP115 proteins and various fragments thereof suitable for use as, e.g., antigens (e.g., for epitope mapping or raising antibodies) and biomarkers for diseases, and for use in therapeutic compositions. Also 5 provided are fusions of GP115 polypeptides to heterologous polypeptides or other moieties (e.g., detectable chemical compounds).

#### **A. GP115 Polypeptides of Particular Sequences**

The invention provides an isolated GP115 polypeptide comprising the amino acid sequence encoded by a full-length GP115 cDNA (SEQ ID NO:2 or a 10 degenerate variant), optionally having one or more conservative amino acid substitutions.

The invention further provides fragments of each of the above-described isolated polypeptides, particularly fragments having at least 5, 6, 8, or 15 amino acids of SEQ ID NO:1. Larger fragments of at least 20, 25, 30, 35, 50, 75, 100, 15 150 or more amino acids are also useful, and at times preferred. The GP115 fragments of the invention may be continuous portions of the native GP115 protein. However, it will be appreciated that knowledge of the GP115 gene and protein sequences as provided herein permits recombining of various domains that are not contiguous in the native GP115 protein.

#### **B. Fusion Proteins And Other Protein Conjugates**

This invention further provides fusions of GP115 polypeptides to heterologous polypeptides. "Fusion" means that the GP115 polypeptide is linearly contiguous to the heterologous polypeptide in a peptide-bonded polymer of amino acids or amino acid analogues. By "heterologous polypeptide" is here intended a 25 polypeptide that does not naturally occur in contiguity with the GP115 fusion partner. As so defined, the fusion can consist entirely of a plurality of fragments of the GP115 protein in altered arrangement; in such a case, any of the GP115 fragments can be considered heterologous to the other GP115 fragments in the fusion protein.

- The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and preferably, at least 15, 20, and 25 amino acids in length. The heterologous sequences can target the GP115 polypeptide to a selected cell by binding to a cell surface receptor, prolong the serum life of the GP115 polypeptide (e.g., an IgG Fc region), make the GP115 polypeptide detectable (e.g., a luciferase or a green fluorescent protein), facilitate purification (e.g., His tag, FLAG, etc.), facilitate secretion of recombinantly expressed proteins (e.g., into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells, through incorporation of secretion signals and/or leader sequences). Other useful protein fusions of the present invention include fusions that permit, e.g., use of the protein of the present invention as bait in a yeast two-hybrid system, or display of the encoded protein on the surface of a phage or cell; fusions to intrinsically delectable proteins, such as fluorescent or light-emitting proteins.
- The proteins and protein fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, ricin, or other biologically deleterious moieties in order to effect specific ablation of cells that bind or take up the proteins of the present invention.
- C. Other Modifications of the Polypeptides**
- The polypeptides of this invention can be composed of natural amino acids linked by native peptide bonds, or can contain any or all of nonnatural amino acid analogues, nonnative bonds, and post-synthetic (post-translational) modifications, either throughout the length of the polypeptide or localized to one or more portions thereof. However, the range of such nonnatural analogues, nonnative inter-residue bonds, or post-synthesis modifications will be limited to those that do not interfere with the biological function of the polypeptide.

Techniques for incorporating non-natural amino acids during solid phase chemical synthesis or by recombinant methods are well established in the art.

For instance, D-enantiomers of natural amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-enantiomers can also be used to confer specific three dimensional conformations on the peptide. Other amino acid analogues

- 5 commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (Kole et al., Biochem. Biophys. Res. Com. 209:817-821 (1995)), and various halogenated phenylalanine derivatives.

- 10 The isolated GP115 polypeptides can also include non-native inter-residue bonds, including bonds that lead to circular and branched forms. The isolated GP115 polypeptides can also include post-translational and post-synthetic modifications, either throughout the length of the protein or localized to one or more portions thereof. For example, when produced by recombinant expression in  
15 eukaryotic cells, the isolated polypeptide can include N-linked and/or O-linked glycosylation, the pattern of which will reflect both the availability of glycosylation sites on the protein sequence and the identity of the host cell. Further modification of glycosylation pattern can be performed enzymatically. As another example, recombinant polypeptides of the invention may also include an initial modified  
20 methionine residue, in some cases resulting from host-mediated processes.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide a labeled polypeptide (e.g., biotin, various chromophores, or fluorophores). The GP115 polypeptides of this invention can also usefully be conjugated to polyethylene glycol (PEG). PEGylation increases the serum half life of proteins administered intravenously for replacement therapy.  
25

#### D. Purification of the Polypeptides

Production of the isolated polypeptides of the present invention can optionally be followed by purification from the producing cells. Producing cells include, without limitation, recombinant cells overexpressing the polypeptides,

naturally occurring cells (e.g., cancer cells) overexpression the polypeptides, or established cancer cell lines overexpressing the polypeptides. If purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tags, such as use of

5       immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immuno-precipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis. Purification of chemically-synthesized peptides can readily be effected, e.g., by HPLC.

10           Accordingly, it is an aspect of the present invention to provide the isolated GP115 proteins of the present invention in pure or substantially pure form. A purified protein of the present invention is an isolated protein, as above described, that is present at a concentration of at least 95%, as measured on a mass basis (w/w) with respect to total protein in a composition. Such purities can often be obtained during  
15           chemical synthesis without further purification, as, e.g., by HPLC. Purified proteins of the present invention can be present at a concentration (measured on a mass basis with respect to total protein in a composition) of 96%, 97%, 98%, and even 99%. The proteins of the present invention can even be present at levels of 99.5%, 99.6%, and even 99.7%, 99.8%, or even 99.9% following purification, as by HPLC.

20           Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents — such as vaccines, or for replacement therapy — the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

25           Thus, the present invention provides the isolated proteins of the present invention in substantially purified form. A "substantially purified protein" of the present invention is an isolated protein, as above described, present at a concentration of at least 70%, measured on a mass basis with respect to total protein in a composition. Usefully, the substantially purified protein is present at a concentration,

measured on a mass basis with respect to total protein in a composition, of at least 75%, 80%, or even at least 85%, 90%, 91%, 92%, 93%, 94%, 94.5% or even at least 94.9%.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

#### E. Exemplary Uses of GP115 Polypeptides

##### 1. Therapeutic Use

The isolated proteins of the present invention can be used as a therapeutic supplement in patients diagnosed to have a specific deficiency in GP115 production or activity.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to its ligand. Thus, such fragments can be used as anti-cancer agents to reduce the activity of GP115.

##### 2. Epitope Mapping

Fragments of at least six contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. See, e.g., Geysen et al., "Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid," Proc. Natl. Acad. Sci. USA 81:3998-4002 (1984) and U.S. Pat. Nos. 4,708,871 and 5,595,915. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least six amino acids of the proteins of the present invention have utility in such a study.

##### 25 3. Immunogens

Fragments of at least eight contiguous amino acids, often at least fifteen contiguous amino acids, have utility as immunogens for raising antibodies that

recognize the proteins of the present invention or as vaccines for GP115-mediated diseases such as cancers.

- The GP115 proteins, fragments, and fusions of the present invention can usefully be attached to a substrate (*supra*). So bound, the new polypeptides can be  
5 used to detect and quantify antibodies, e.g., in serum, that bind specifically to the immobilized protein of the present invention.

#### IV. ANTIBODIES AND ANTIBODY-PRODUCING CELLS

##### A. General Consideration

- The invention provides antibodies that bind specifically to the new  
10 GP115 polypeptides. The antibodies can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such polypeptides, either as present on the polypeptide in its native conformation or, in some cases, as present on the polypeptides as denatured, as, e.g., by solubilization in SDS. In some embodiments, the invention provides antibodies, both polyclonal and monoclonal, that bind  
15 specifically to a polypeptide having an amino acid sequence presented in SEQ ID NO:1.

An antibody of this invention refers to a full antibody, e.g., an antibody comprising two heavy chains and two light chains, or to an antigen-binding fragment of a full antibody. Such fragments include, but are not limited to, those produced by  
20 digestion with various proteases, those produced by chemical cleavage and/or chemical dissociation, and those produced recombinantly, so long as the fragment remains capable of specific binding to an antigen. Among these fragments are Fab, Fab', F(ab')<sub>2</sub> and single chain Fv (scFv) fragments.

An antibody of this invention can be a murine or hamster antibody or a  
25 homolog thereof, or a fully human antibody. An antibody of this invention can also be a humanized antibody, a chimeric antibody, an antibody fusion, an diabody, an intrabody, or a single-chained antibody. An antibody of this invention can be of any isotype and subtype, for example, IgA (e.g., IgA1 and IgA2), IgG (e.g., IgG1, IgG2,

IgG3 and IgG4), IgE, IgD, IgM, wherein the light chains of the immunoglobulin may be of type kappa or lambda. While the useful antibodies are generally monoclonal, polyclonal antibodies from mice, rabbits, turkeys, or sheep may also be used.

Typically, the affinity or avidity of an antibody (or antibody multimer, 5 as in the case of an IgM pentamer) of the present invention will be at least about  $1 \times 10^{-6}$  molar (M), typically at least about  $5 \times 10^{-7}$  M, usefully at least about  $1 \times 10^{-7}$  M, with affinities and avidities of at least  $1 \times 10^{-8}$  M,  $5 \times 10^{-9}$  M, and  $1 \times 10^{-10}$  M proving especially useful.

#### B. Moieties Conjugated to the Antibodies

The antibodies are useful in a variety of *in vitro* immunoassays, such as 10 Western blotting and ELISA, in isolating and purifying GP115 proteins (e.g., by immunoprecipitation, immunoaffinity chromatography, or magnetic bead-mediated purification). The antibodies are also useful as modulators (i.e., antagonists or agonists) of a GP115 protein *in vivo* to modulate the protein's interaction with its 15 natural ligand. The antibodies can also be used to conjugate to cytotoxic reagents for site-specific delivery.

The new antibodies can be variously associated with moieties appropriate for their uses. By way of example, when the antibodies are used for 20 immunohistochemical staining of tissue samples, the moieties can be an enzyme that catalyzes production and local deposition of a detectable product. Exemplary enzymes are alkaline phosphatase,  $\beta$ -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. The antibodies can also be labeled using colloidal gold. When the antibodies are used, e.g., for flow cytometric detection and scanning 25 laser cytometric detection, they can be labeled with fluorophores. For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin. When the antibodies of the present invention are used, e.g., for Western blotting, they can usefully be labeled with radioisotopes. When the antibodies are to be used for *in vivo* diagnoses, they can

be rendered detectable by conjugation to MRI contrast agents, such as radioisotopic labeling or gadolinium diethylenetriaminepentaacetic acid (DTPA).

The antibodies of this invention can also be conjugated to biologically deleterious agents so as to direct the agents to a tumor site. By way of example, the 5 antibody is conjugated to *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. See, e.g., Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, Vol 166), Humana Press (2000) (ISBN:0896037754); and Frankel et al. (eds.), Clinical Applications of Immunotoxins, Springer-Verlag New York, Incorporated (1998) (ISBN:3540640975). Small 10 molecule toxins such as calicheamycin or chemotherapeutic agents can also be delivered via chemical conjugation to the antibodies. The antibodies may also be used to deliver DNA to the tumor site as gene therapy to inhibit or otherwise modify the behavior of the tumor (e.g., to deliver an antisense reagent to the GP115 oncogene).

For some uses, the antibodies can be bound to a substrate via a linker 15 moiety. For example, the antibodies can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for immunoaffinity chromatography. The antibodies can also be attached to paramagnetic microspheres, by, e.g., biotin-streptavidin interaction, which microsphere can then be used for isolation of cells that express or display the proteins of the present invention. As 20 another example, the antibodies can also be attached to the surface of a microtiter plate for ELISA.

## V. PHARMACEUTICAL COMPOSITIONS

As a protein involved in tumor genesis, development (including maintenance), GP115 will be a suitable therapeutic target for treating neoplasia, 25 hyperplasia, malignant cancers, or any other hyperproliferative conditions.

The invention accordingly provides pharmaceutical compositions comprising nucleic acids, proteins, and antibodies of the present invention, as well as mimetics, agonists, antagonists, or modulators of GP115 activity, and methods of

using them to prevent or treat (i.e., ameliorate, mitigate, alleviate, slow, or inhibit) tumor growth, angiogenesis, metastasis or otherwise inappropriate cell proliferation.

- A composition of the invention typically contains from about 0.1 to 90% by weight (such as 1 to 20% or 1 to 10%) of a therapeutic agent of the invention
- 5 in a pharmaceutically accepted carrier. Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as corn starch, gelatin, lactose, acacia, sucrose, microcrystalline cellulose, kaolin, mannitol, dicalcium phosphate, calcium carbonate, sodium chloride, or alginic acid. Disintegrators that can be used include, without limitation, microcrystalline cellulose, corn starch, sodium
- 10 starch glycolate, and alginic acid. Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone(Povidone<sup>TM</sup>), hydroxypropyl methylcellulose, sucrose, starch, and ethylcellulose. Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.
- 15 Liquid formulations of the compositions for oral administration prepared in water or other aqueous vehicles can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active
- 20 compound(s), wetting agents, sweeteners, and coloring and flavoring agents. Various liquid and powder formulations can be prepared by conventional methods for inhalation into the lungs of the mammal to be treated.

- Injectable formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl
- 25 carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like). Physiologically acceptable excipients can include, for example, 5% dextrose, 0.9% saline, Ringer's solution or other suitable excipients. Intramuscular preparations, e.g., a sterile formulation of a suitable soluble salt form of the compounds, can be dissolved and administered in a pharmaceutical excipient such

as Water-for-Injection, 0.9% saline, or 5% glucose solution. A suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (e.g., ethyl oleate).

- 5           A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, e.g., 5 to 10%, in a carrier such as a pharmaceutical cream base. Various formulations for topical use include drops, tinctures, lotions, creams, solutions, and ointments containing the active ingredient and various supports and vehicles. The optimal percentage of the  
10 therapeutic agent in each pharmaceutical formulation varies according to the formulation itself and the therapeutic effect desired in the specific pathologies and correlated therapeutic regimens.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th  
15 ed., Lippincott, Williams & Wilkins (2000) (ISBN: 0683306472); Ansel et al., Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins Publishers (1999) (ISBN: 0683305727); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000) (ISBN: 091733096X), the disclosures of which are incorporated herein by  
20 reference in their entirieties. Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) to the patient.

Typically, the pharmaceutical formulation will be administered to the patient by applying to the skin of the patient a transdermal patch containing the  
25 pharmaceutical formulation, and leaving the patch in contact with the patient's skin (generally for 1 to 5 hours per patch). Other transdermal routes of administration (e.g., through use of a topically applied cream, ointment, or the like) can be used by applying conventional techniques. The pharmaceutical formulation(s) can also be administered via other conventional routes (e.g., parenteral, subcutaneous,

intrapulmonary, transmucosal, intraperitoneal, intrauterine, sublingual, intrathecal, or intramuscular routes) by using standard methods. In addition, the pharmaceutical formulations can be administered to the patient via injectable depot routes of administration such as by using 1-, 3-, or 6-month depot injectable or biodegradable

5 materials and methods.

Regardless of the route of administration, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (e.g., 1mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired  
10 daily dose. The effectiveness of the method of treatment can be assessed by monitoring the patient for known signs or symptoms of a disorder.

The pharmaceutical compositions of the invention may be included in a container, package or dispenser alone or as part of a kit with labels and instructions for administration. These compositions can also be used in combination with other cancer  
15 therapies involving, e.g., radiation, photosensitizing compounds, anti-neoplastic agents and immunotoxics.

## VI. GP115-RELATED ANIMALS

### A. General Consideration

This invention provides also nonhuman transgenic animals whose somatic and germ cells contain a GP115-coding nucleic acid of this invention (including both heterozygotes and homozygotes). Such animals can be used to study the effect of the GP115 gene on tumorigenicity and tumor development, to study the role of GP115 on normal tissue development and differentiation, to identify via array CGH regions of the genome whose amplification or deletion is correlated with GP115  
20 status, and to screen for and establish toxicity profiles of anti-cancer drugs. This invention also provides transgenic animals with targeted disruption of one or both copies of the endogenous GP115 gene. Also included are chimeric animals that can be used to generate the transgenic animals. As used herein, a "transgenic animal" is a  
25 non-human animal, preferably a mammal, more preferably a cow, goat, sheep, or

rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, dogs, chickens, amphibians, etc.

#### B. Inducible Cancer Model

5 In some embodiments, this invention provides an inducible cancer model to study tumor biology and to screen for anti-cancer drugs. In one embodiment, a transgenic animal has a genome comprising: (a) an expression construct comprising a GP115 gene linked operably to an inducible promoter, and (b) a genetic mutation that causes said transgenic mouse to have greater susceptibility to cancer than a mouse  
10 not comprising said genetic mutation, wherein expression of said GP115 gene leads to formation of cancer in said transgenic mouse, and wherein said cancer regresses in said transgenic mouse when expression of said GP115 gene is reduced. Mutations that render the animal more susceptible to cancer include disabling mutations in a tumor suppressor gene (e.g., INK4a), disabling mutations in a DNA repair gene (e.g.,  
15 MSH2), and activating mutations in an oncogene (e.g., *myc* and *ras*). Such testing also can be carried out in cells (e.g., human cells) that are engineered to contain an inducible oncogene and endowed with tumorigenic capacity by the presence of an appropriate combination of oncogenes, tumor suppressor genes, and/or telomerase.

In one particular embodiment, the animal's genome comprises (i) a first  
20 expression construct containing a gene encoding a reverse tetracycline transactivator operably linked to a promoter, such as any tissue or cell type-specific promoter or any general promoter, and (ii) a second expression construct containing the GP115 gene operably linked to a promoter that is regulated by the reverse tetracycline transactivator and tetracycline (or a tetracycline analogue, for example, doxycycline),  
25 and (b) observing the mammal in the presence and absence of the tetracycline (or analogue thereof) for the development, maintenance, or progression of a tumor that is affected by the presence or absence of the tetracycline (or analogue thereof). Of course, other inducible systems such as those described can also be employed.

This animal model can be used to determine the efficacy of a candidate compound in preventing or treating cancer. This method involves administering to the animal a candidate compound and observing the effect of the compound on tumor development, maintenance, angiogenesis and/or progression in the animal. Regression and/or reduction of tumor size in the presence of the compound is indicative of the effectiveness of the compound. Similarly, the effect of a candidate compound on the level of GP115 mRNA, protein, or activity in the animal or cell lines derived thereof (or cell lines transfected with the gene) can be used to identify the candidate as an agonist or antagonist. Also, the ability to compare the effect of a test compound to that of genetically switching off the inducible oncogene in this system allows the identification of surrogate markers that are predictive of the clinical response to the compound. Lastly, the inducible model can be used to determine whether a compound can eradicate minimal residual tumor. Normally in the inducible model, a tumor regresses when the GP115 gene is switched from "on" to "off" via the inducible promoter. But if a compound can eradicate minimal residual tumor, switching the gene back on after administration of the compound will not bring back the tumor.

The animal model can also be used to identify other cancer-related elements. To do this, a detailed expression profile of gene expression in tumors undergoing regression or regrowth due to the inactivation or activation of the GP115 transgene is established. Techniques used to establish the profile include the use of suppression subtraction (in cell culture), differential display, proteomic analysis, serial analysis of gene expression (SAGE), and expression/transcription profiling using cDNA and/or oligonucleotide microarrays. Then, comparisons of expression profiles at different stages of cancer development can be performed to identify genes whose expression patterns are altered.

This animal model can also be used to identify molecular surrogates of GP115 in another manner. To do this, the expression of GP115 gene is eliminated by null mutation, and another round of MaSS screening is performed using retroviral integration, cDNA complementation, or the genetic suppressor elements (GSE) method.

Genes whose activation results in transformation of the cells are likely to be in a related tumorigenic pathway as GP115.

The animal model can also be used to identify surrogate biomarkers for diagnosis or for following disease progression in patients. The biomarkers can be  
5 identified based on the differences between the expression profiles of the "on" and "off" states in the animal model. Blood or urine samples from the animal can be tested with ELISAs or other assays to determine which biomarkers are released from the tumor into circulation during tumor genesis, maintenance, or regression (when GP115 is turned off). These biomarkers are particularly useful clinically in following  
10 disease progression post anti-GP115 therapy.

## VII. CANCER DIAGNOSTICS

Since GP115 activity is up-regulated in tumor cells, one can use GP115 as a marker in diagnosing cancer or any other abnormal hyperplasia conditions. To do this, the nucleic acid probes or antibodies of this invention are used to quantify the  
15 expression level of GP115 in a tissue sample, wherein an increase in that level relative to control is indicative of cancerous, neoplastic, or hyperplastic pathology of the tissue sample. Routine techniques include RT-PCR, ribonuclease protection assays, *in situ* hybridization, Northern blot analysis, FISH, CGH, array CGH, SKY, and immunohistochemistry.

Because up-regulation of GP115 is generally associated with a malignant state, a GP115 protein or fragments thereof may be found to be elevated in a tissue sample (e.g., blood or urine) of cancer patients relative to that of normal individuals. This elevation can be detected by, e.g., specific ELISAs, radioimmunoassays, or protein chip assays. Such tests may not only be useful for  
25 diagnosis of GP115-related diseases such as cancers, but also for monitoring the progress of therapy using GP115 inhibitors.

## VIII. EXAMPLES

The following examples are intended to illustrate the methods and materials of the present invention. Suitable modifications and adaptations of the described conditions and parameters normally encountered in the art that are obvious to those skilled in the art are within the spirit and scope of the present invention.

### Example 1: MaSS Screening Identification of the Gene

This example describes the procedures for identifying the GP115 gene by MaSS screening.

#### a. Retroviral Infection of Tumor Cells

Mo-MuLV producer cell line TMJ (NIH3T3 based cell line) was plated to the required number of plates (100 mm). These cells were cultured and maintained in RPMI media with 10% FBS. For viral production, TMJ cells were fed with 4-5 ml of fresh culture media, and culture supernatant was harvested 8-12 hours later. The supernatant was filtered through a 0.45  $\mu$ M filter.

Meanwhile, doxycycline-dependent, RAS-induced melanoma cells such as R545 cells were maintained in RPMI media with 10% fetal calf serum in the presence of doxycycline (2  $\mu$ g/ml). At approximately 18-24 hrs after plating or when the plates were 70-80% confluent, the R545 cells were infected with the filtered viral supernatant in the presence of polybrene (6-8  $\mu$ g/ml). From this point on, the R545 cells were maintained in the absence of doxycycline.

Eighteen hours after infection, infected R545 cells were trypsinized, rinsed and resuspended in Hanks' Balanced Salt Solution. Cell suspensions were kept on ice and the handling time after trypsinization was kept to a minimum. About 1 X  $10^6$  cells were injected onto the flank of SCID mice fed with water without doxycycline. The animals were observed for tumor development. Control animals were similarly injected with 1 X  $10^6$  uninfected R545 cells. Tumors typically developed after approximately 21 days. Tumors were harvested and tumor tissues were immediately snap-frozen in LN<sub>2</sub>.

**b. Inverted Polymerase Chain Reaction**

DNA was isolated from tumor tissues using the PUREGENE DNA isolation kit. 10 µg of genomic DNA was digested to completion with either BamHI or SacII and the reaction was terminated by incubation at 65°C for 20 minutes. The 5 digested samples were self-ligated in a diluted 600 µl reaction volume using 4000 U of high concentration T4 Ligase (NEB, Cat. # M0202M). The ligation was performed overnight to 24 hrs at 16°C. The ligated DNA was precipitated with ethanol and dissolved in 40 µl of sterile water. The ligated DNA was then serially diluted to 1:10 and 1:100 ratios and subjected to inverted polymerase chain reaction (IPCR).

10 The PCR reaction mix had a total volume of 50 µl and contained 1µl of the ligated DNA, 25 nmol of each dNTP, 10 pmol each of the forward and reverse primers, 1 X Buffer 2, and 2.5 U of Enzyme Mix in the EXPAND Long Template PCR System (Roche). Amplification was performed as follows: 92°C for 2 min, then 10 cycles of (92°C for 10 sec, 63° for 30 sec, 68°C for 15 min), then 20 cycles of 15 (92°C for 10 sec, 63°C for 30 sec, 68°C for 15 min, and a 20 sec auto extension), and a final extension step at 68°C for 30 min (TETRAD Thermocycler, MJ Research). The primer sets used in IPCR (all of them targeting the retroviral LTRs) were:

S5'1F: GAGGCCACCTCCACTTCTGAGAT (SEQ ID NO:10);

S5'1R: CTCTGTCGCCATCTCCGTCAGA (SEQ ID NO:11);

20 S5'2F: CAUCAUCAUCCTGGCCCTCTCCCATAGTGT (SEQ ID NO:12);

S5'2R: CUACUACUACUAGGCAGTTACTGCAGTTAGCTGGCT (SEQ ID NO:13);

S3'1F: GGCTGCCATGCACGATGACCTT (SEQ ID NO:14);

25 S3'1R: CGGCCAGTACTGCAACTGACCAT (SEQ ID NO:15);

S3'2F: CUACUACUACUAGGGAGGGTCTCCTCAGAGTGATT (SEQ ID NO:16);

S3'2R: CAUCAUCAUCAUGGAAAGCCCGAGAGGTGGT (SEQ ID NO:17);

B3'1R: CGGGAAGGTGGTCGTCGGTCT (SEQ ID NO:18); and

B3'2R: CAUCAUCAUCAUGGGGCCCGAGTCTGTAATTT (SEQ ID 5 NO:19).

For BamHI 5' cloning, primary PCR was done by using S5'1F and S5'1R followed by nested PCR using S5'2F and S5'2R. For BamHI 3' cloning, primary PCR was done by using S3'1F and B3'1R followed by nested PCR using S3'2F and B3'2R. For SacII 5' cloning, primary PCR was done by using S5'1F and 10 S5'1R followed by nested PCR using S5'2F and S5'2R. For SacII 3' cloning, primary PCR was done by using S3'1F and S3'1R followed by nested PCR using S3'2F and S3'2R.

The PCR products were resolved on 1% agarose gel. Individual bands were excised and purified using the QIAGEN Gel Extraction kit. The purified DNA 15 was dissolved in 30 µl of sterile water and subjected to direct DNA sequencing.

#### c. Identification of Candidate Genes

The site of retroviral integration into the mouse genome was mapped for all IPCR sequences as follows. Retroviral leader sequences were trimmed from the raw sequences of IPCR products, and homology searches for the trimmed 20 sequences were performed in the NCBI MGSCV3 database by using the BLAST software program. BLAST hits were analyzed and recurrent sites of integration in multiple mouse tumors were identified. Recurrence was defined as 2 or more integrations within a 10 kb region. To identify genes whose expression was affected by the retroviral integration, NCBI MapView was used to identify the site of each 25 recurrent retroviral integration onto the mouse genome. Genes immediately neighboring the site were identified by using the MGSCV3 Gene map. These genes were defined as candidate cancer-related genes because in the vast majority of cases, MuLV integration affects the most proximal genes. When the integration occurred

within a gene, that gene was deemed the best candidate as the target for the effects of retroviral integration.

#### **Example 2: Expression in Human Tumors and Tumor Cell Lines**

This example describes the protocols for expressing the candidate gene  
5 identified above in human cancer cells.

Primer pairs for each human gene were designed as described for the mouse gene. Expression of each candidate gene was assessed in a panel of 31 human cancer cell lines and 47 human primary tumors by using real-time reverse transcription PCR. The forward and reverse primers were 5'-ggcagcttcattggtctc-3' (SEQ ID NO:20) and 5'-caacgttagccgaaagtca-3' (SEQ ID NO:21), respectively. RNA was prepared from the cells using QIAGEN RNEASY mini-prep kits and QIAGEN RNEASY maxi-prep columns. RNA preparations were treated with DNase during column purification according to manufacturer's instructions. Expression of each gene was determined in triplicate for all tumors and cell lines using SYBR green-based real-time PCR. To do this, 2X SYBR green PCR master mix (ABI) was mixed with the MULTISCRIBE reverse transcriptase (ABI) and RNase inhibitor (ABI). Forward and reverse primers were added at ratios previously optimized for each gene using control human reference RNA (Stratagene). 50 ng of RNA template was used per reaction, and the reactions were performed in a total volume of 20 µl. Real-time quantification was performed using the ABI 7900HT and SDS2.0 software. RNA loading was normalized for β-actin and 18S rRNA. RNA quantity was determined relative to human universal reference RNA (Stratagene) to permit run-to-run comparisons.

#### **Example 3: RNAi Inhibition of Human Cancer Cell Lines**

This example describes the protocols used to inhibit the expression of  
25 the candidate gene in human cancer cells by RNA interference (RNAi).

##### **a. Generation of siRNA**

Double-stranded siRNA oligonucleotides were designed using the OligoEngine siRNA design tool (<http://www.oligoengine.com>). HPLC purified siRNA oligonucleotides were incubated at 95°C for 1 min, 37°C for 1 hr, and room temperature for 30 min. siRNAs were stored at -20°C prior to use.

5 Human cancer cell lines were transfected with siRNA using the Oligofectamine reagent from Invitrogen. Transfected cells were allowed to grow for at least 24 hours. The cells were then trypsinized and reseeded for growth curve analysis or growth in soft agar.

**b. Growth Curves**

10 To create growth curves, early passage cell lines were seeded at  $2 \times 10^4$  cells per well in 12-well plates. At 6 (Day 0), 24 (Day 1), 48 (Day 2), 72 (Day 3), 96 (Day 4) and 120 hours (Day 5) after plating, duplicate plates of cells were washed, fixed in 10% buffered formalin, and stained with crystal violet for 30 min at room temperature. Stained cells were washed in double distilled water, and stain was  
15 extracted using 10% acetic acid. Absorbance of the extracted stain was read at 590 nm. The mean absorbance per well of duplicate cultures was determined.

**c. Soft Agar Assay**

One day prior to seeding cells for assay, 6-well plates containing bottom agar were prepared. The bottom agar was made of 0.7% agarose (SeaKem  
20 GTG agarose) and 1X DME plus 10% fetal bovine serum. The next day, cells were seeded into each well by adding 5ml of a mixture containing  $1 \times 10^4$  cells in 0.32% agarose and 1 X DME plus 10% FBS. The cell mixture was allowed to solidify at room temperature for 30 min. The plates were then incubated at 37°C in a 5% CO<sub>2</sub> atmosphere for 1 to 14 days. Colony formation was analyzed and photographs taken  
25 at various time points.

**Example 4: MEF Transformation**

This example describes a protocol for using the candidate gene to transform mouse embryonic fibroblasts.

MEFs are isolated from 100 individual 13.5-day-old embryos. The isolated MEFs are pooled and grown in DMEM supplemented with 10% FBS, penicillin and streptomycin. Pooled early-passage (passages 4 to 6) p16<sup>Ink4a</sup>/p19<sup>Arf</sup> -/- mouse embryonic fibroblasts are transfected with (1) pKO-Myc; (2) pKO-Myc and 5 pT24-RasV12; (3) pKO-Myc and GP115 cDNA; or (4) pKO-Myc, pT24-RasV12, and GP115 cDNA. All transfections are done in duplicate cultures ( $8 \times 10^5$  cells) using Lipofectamine Plus in Optimem (Gibco) without serum or antibiotics. Cultures are split 1:3 a day after transfection. Foci are counted on day 12.

**Example 5: RNAi on Human or Mouse model Cancer Cell Lines in SCIDs**

10 Human cancer cell lines expressing high levels of the candidate, or cell lines established from the inducible mouse cancer model described above, are transfected with a vector encoding a short hairpin RNA homologous to the candidate cancer-related gene. Expression of the RNA is placed under the control of an inducible U6 promoter. Stable cell lines are established.  $5 \times 10^5$  cells are injected 15 subcutaneously into the flank of 6 week old female inbred SCID mice. Tumor formation is observed visually. For cell lines derived from the inducible mouse model, tumor formation is induced by doxycycline. Expression of the RNAi is induced once tumors were visually identified. In the case of mouse model-derived tumor cells, the mice are fed with doxycycline. Tumor regression is followed using 20 calipers to measure the shrinking tumor diameter.

**Example 6: Tumor Formation in SCIDs**

Tumor cells derived from the tumor suppressor null (INK4arf -/-) doxycycline-inducible oncogene mouse model are infected with retrovirus encoding the candidate gene under the control of an IPTG-inducible promoter. Stable cell lines 25 are established.  $10^6$  cell are injected subcutaneously into the flank of 6 week old female inbred SCID mice. Mice are fed with doxycycline for 7-12 days. 24 hours prior to doxycycline withdrawal, mice are fed IPTG. IPTG feeding is maintained after withdrawal of doxycycline and tumor regression is monitored using calipers.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can

5 also be used in the practice or testing of the present invention. All publications and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. The materials, methods, and examples are illustrative only and not intended to be limiting. Throughout this specification and claims, the word "comprise," or variations such as

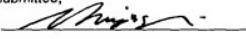
10 "comprises" or "comprising" will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

015711 S.P.T.O.  
02/27/03Please type a plus sign (+) inside this box → PTO/SB/16 (8-00)  
Approved for use through 10/31/2002 GMP 065-1-0032U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE  
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.**PROVISIONAL APPLICATION FOR PATENT COVER SHEET**

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

INVENTOR(S)					
Given name (first and middle [if any])	Family Name or Surname	Residence (city and either State or Foreign Country)			
Ronan C. Kuruppuhan David	O'Hagan Kannan Bailey	Brookline, MA Cambridge, MA Brighton, MA			
<input type="checkbox"/> Additional inventors are being named on the _____ separately numbered sheets attached hereto					
TITLE OF THE INVENTION (280 characters max)					
GP115: METHODS AND COMPOSITIONS FOR TREATING CANCER					
Direct all correspondence to: <b>CORRESPONDENCE ADDRESS</b>					
<input type="checkbox"/> Customer Number <b>1473</b> → <input type="checkbox"/> Place Customer Number Bar Code Label here OR <input type="checkbox"/> Type Customer Number here					
<input checked="" type="checkbox"/> Firm or Individual Name	Z. Ying Li				
Address	c/o Fish & Neave				
Address	1251 Avenue of the Americas				
City	New York	State	N.Y.	ZIP	10020
Country	USA	Telephone	212-596-9000	Fax	212-596-9090
ENCLOSED APPLICATION PARTS (check all that apply)					
<input checked="" type="checkbox"/> Specification	Number of Pages	49	<input type="checkbox"/> CD(s), Number		
<input checked="" type="checkbox"/> Drawing(s)	Number of Sheets	0	<input type="checkbox"/> Other (specify)		
<input checked="" type="checkbox"/> Application Data Sheet. See 37 CFR 1.76					
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT					
<input checked="" type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27. <input checked="" type="checkbox"/> A check or money order is enclosed to cover the filing fees <input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: <input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.				<b>06-1075</b>	<b>FILING FEE AMOUNT (\$)</b> <b>\$80.00</b>
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government. <input checked="" type="checkbox"/> No. <input type="checkbox"/> Yes, the name of the U.S. Government agency and the Government contract number are: _____					

Respectfully submitted,

SIGNATURE TYPED or PRINTED NAME Z. Ying LiTELEPHONE 212-596-9106

Date: December 16, 2002

REGISTRATION NO.  
(if appropriate)  
Docket Number:**42,800****GEN-25PROV****USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT**

This collection of information is required by 37 CFR 1.51. The information is used by the public to file (and by the PTO to process) a provisional application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the complete provisional application to the PTO. The burden may be reduced by providing updated information each time a communication is made with the PTO. The PTO may provide suggestions for reducing this burden. These suggestions, if accepted, may result in changes to future versions of this form. You may send comments on the burden estimate or suggestions for reducing this burden to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

15535  
60/198193  
06/27/03

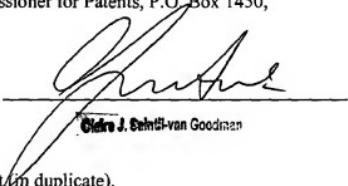
Applicants : Ronan C. O'Hagan, et al.  
Application No. : To Be Assigned Confirmation No.:  
Filed : Herewith  
For : GP115: METHODS AND COMPOSITIONS FOR  
TREATING CANCER

New York, New York 10020  
August 27, 2003

Hon. Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

"Express Mail" mailing label number EV045438026US  
Date of Deposit August 27, 2003

I hereby certify that this paper/fee is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to Mail Stop Provisional Application to the Hon. Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

  
Charles J. Schmid-van Goodman

Encl:

- Provisional Patent Application Cover Sheet (in duplicate).
- Provisional Patent Application (49 pages).
- Print EFS Data Sheet.
- Check in the amount of \$80.00